# Cognitive Information Retrieval Based on Ontological Model of Knowledge Representation

N.V. Maksimov, O.L. Golitsina, K.V. Monankov

National Research Nuclear University MEPhI, Moscow, Russian Federation

Email contact of corresponding author: nv-maks@yandex.ru, kmonankov@yandex.ru

**Abstract**

Information retrieval technologies in databases with full-text semantic indexing are considered. The information retrieval process is considered as a cognitive-oriented process. А document semantic image is presented as an ontology. An ontology is defined as the set of three interconnected systems (functional, conceptual and terminological), on which the operation of comparing elements of different systems at the level of signs is defined. A functional system (a tasks system, objects, processes, subject area properties) represents objects and situational relationships between them in the context of the target activity. The conceptual system objects are stable concepts, and the set of relationships is limited to generic and associative relationships. The terminological system reflects the properties of a natural language at the level of signs – terms for which relationships of equivalence and lexicographic inclusion, as well as linguistic relationships, are specified. Document semantics is represented as an ontology graph (knowledge graph). Such graph in vertices contains named entities (concepts, names, values, etc.), and in arcs – typed relationships, extracted from the text also taking into account their location. This made it possible to build up mechanisms (algorithms) for a new type of IR – searching semantic dependencies and neighborhoods within the document text or a documents selection. Such cognitive information retrieval is considered as a search for a path on an ontology graph dynamically formed based on found documents ontological images or their fragments. A Cognitive Subject Tree is used to fix and manage search directions. A Cognitive Subject Tree is a hierarchically ordered structure of a personified representation of a subject area (project, task). Such a structure integrally identifies tasks / knowledge, since each structural and semantic component (node) of the tree includes not only keywords, but also relevant documents texts, query expressions, indexes of classifiers, etc. Example of a search on knowledge graphs on the topic “Proryv Project” is considered.

## INTRODUCTION

The well-known metaphor “Standing on the shoulders of giants” tells us that any new knowledge is somehow based on existing knowledge. Modern effective search and analysis tools of an existing knowledge in the context of vast amounts of information are needed for an effective synthesis of new knowledge today. At the same time, there is a clear trend of a natural merging of technologies of main activity (MAc) and information activity (IAc).

In the conditions of integrated digitalization of the scientific, industrial, social sphere, with the accompanying "movement" of data, knowledge and the processes of their creation and use into a distributed computing environment, information search should be considered as one of the tools of cognition, where the search tasks will be not only finding topic publications, but also analyzing the state of the subject area, and also, information modeling of the synthesized result and, to some extent, the process of cognition itself. Such search could be called informational cognition – a mutually coordinated search for situationally relevant fragments of knowledge outside and inside the subject's consciousness.

In the field of IRS there is not acknowledged definition of the cognitive search concept today. For example, in [1] a cognitive search is presented as a tool focused on solving the following tasks: removal of synonymy and polysemy, analysis of abbreviations, morphological analysis of text, removal of ambiguity, calculations on ontologies. In [2] cognitive search considers as tools of a search index enriching through the artificial intelligence methods use such as pattern recognition – extracting text from images, videos, voice recordings, and other unstructured data sources, as well as the use of deep natural language processing. In [3] author discusses the tools of interactive interaction with information resources containing big data, and provides methods for supporting query formulation (hint, error handling), facets formation, ranking of results, highlighting query terms and geodata referencing. Note, however, that the nature of functions attributed to cognitive search in most articles rather corresponds to the conceptual search (search by keywords). In [4] cognitive search is defined as a “tools and technologies to support self-service extraction of information and new insights from large repositories of unstructured and structured data that resides in multiple sources such as file systems, databases, streams, APIs, other platforms, and applications” [4]. Semantic search in according to [5] denotes search with meaning. This “meaning” can refer to various parts of the search process: understanding the query (instead of just finding matches of its components in the data), understanding the data (instead of just searching it for such matches), or representing knowledge in a way suitable for meaningful retrieval. In [6] “Knowledge Graph” is defined as a way of structuring and representing text encoding scientific knowledge. In [7, 8] we suggest a cognitive-similar information retrieval as a search for a path or neighborhood on a multi-meta-hypergraph of an ontology, dynamically formed on the base of ontological images of found documents or their fragments.

In this paper, cognitive information retrieval is defined as: (a) an integrated technology designed for information support of cognitive processes; (b) a system and technology with operands/operations and structures/processes similar to cognitive ones.

“Cognitive” here indicates that an ultimate goal of the generalized human-machine system “cognition – information retrieval” is an ordering and synthesis of knowledge, i.e. a cognitive information retrieval system (unlike classical bibliographic or full-text IRSs) should have functions and declarative tools that synchronize and informationally coordinate objects and processes in a mind and in a computing environment. In particular, the “cognitive” property of the system in such interaction can be ensured by using search technologies similar to functions of consciousness, such as abstraction, generalization, classification, analysis, synthesis, etc.

“Integrated” technology here indicates a steady trend of technological and semantic integration in a general computing environment of MAc and IAc processes and objects in an activity “quantization” mode. Queries that arise in a process of solving a MAc practical problem will be related to a minimal, but significant for decision-making semantic “blocks”, and answers (documents fragments, as semantic elements, somehow useful for solving a pragmatic problem) will may provoke new questions. And in this dialogue a conceptual system will develop along the way. This is fundamentally different from the “ideology” of classical information retrieval, built on the query-answer scheme, which assumes that a formally complete output is formed according to a semantically complete query expression (in so-called “batch mode”).

Ontologies and their graph forms (knowledge graphs) use in IRSs today makes it possible to effectively implement deep semantic search (with reference to a context) with a possibility of technologically closely linking it with a MAc processes. Indexing tools can break a text into heterogeneous identifiable fragments (word, phrase, sentence, paragraph, etc.). Moreover, according to one text fragment, for example, representing a compound concept, the concepts that form it can still be separated. This will lead to an increase in a number of entry points into a search space, which, on the one hand, will increase redundancy, but on the other hand, increase search accuracy, and will also enable a user to choose concepts that are most adequate to a real need.

However, here we should also remember the condition of visibility: even a relatively small text of a few Kb generates a graph, elements of which (and, accordingly, their labels) will be indistinguishable when visualized in usual sizes for a person (A4-A3 format). Therefore, it is necessary to create such a semantic search tool that would allow not only to search formally relevant documents, but also to extract semantically coherent text fragments from them, while forming an operating space of such a size that would be acceptable for visual perception and processing.

Deep semantic search technologies based on ontological and graph approaches are considered in this paper. Examples of semantic scaling operation (namely aspect projection and enlargement as objects aggregation - part-whole relationship) as a part of search operations on knowledge graphs are given. In addition, Cognitive Subject Tree as a tools of integral presentation and analysis of search results and knowledge structure is considered. These tools allow conducting a comparative analysis of information components (development branches) of a project. The examples were prepared in the information-analytical system xIRBIS environment [9] using the INIS thesaurus and taking into account the experience of building the taxonomy of SVBR-100 fast reactors.

## Information retrieval on knowledge graphs

Cognitive oriented information retrieval process, in general, is iterative and not one-act, because here, at query expression adapting, objects of three spaces should be semantically linked: mental, operational (interface), machine. Besides, cognitive processes are characterized by uncertainty, not for result only, but also, for example, how and why to obtain it. That is, in the cognition process (and information retrieval, which is an integral cognition part) an object, a method, and a goal may change. Accordingly, in an information retrieval process it will be necessary to “reformulate” information needs hierarchy forms itself: both expressed and formalized, and conscious and real, since none of them can be considered predetermined and precisely set. Such retrieval for a “cognition path” is an interactive (with alternation of MAc and IAc), sequential, dependent on a previous result, some “deducible fact” formation on facts chaotic set by constructing a valence concept neighborhood and choosing an arc whose relation class corresponds to a functional relationship character predetermined by a problem nature.

Semantic information retrieval, oriented to cognitive processes support, is reduced to a successive performance of document selection (information retrieval classical task of document selection by formal relevance) and deep semantic search itself – a found documents content analysis using document(s) ontology graph as a text and concept navigation tool. That is, the traditional information retrieval technology is supplemented by stages of interactive analysis and search for subgraphs corresponding to a need essence (and not only to a search query expression) in a visualized graph and results evaluation of relevant text fragments combining (corresponding to selected vertices) as possible new knowledge elements.

The concept “deep semantic search” distinctive feature is not only that a meaning is represented by all interrelated facts presented in text (triplets – sufficiently precise entities connected by classified relations), but also that a separate concept meaning (or some semantically interrelated set of them) will be “visible” directly in same space where the search is performed. At the same time, we accept that the specific meaning of any concept is determined by the context – that part of its environment that corresponds, perhaps not completely, to the subject of the current information need. That is, understanding (meaning formation in a user's mind) is a mental image construction from elements (concepts and connections) presented in an operational (and in this case, visual) space, and from available knowledge elements. That is, essential is that the actual “construction” itself is ordered and follows explicitly or implicitly some “assembly” scheme that correlates both with a task cognitive state and with user's perception peculiarities. It is assumed that a new knowledge unit can be constructed from documents fragments found by a query. In this context, the elementary fact represented by a triplet can be considered as a kind of particular meaning marker contained in a separate text fragment, and a triplet connection with a corresponding fragment, preserved during indexing, makes it possible to go directly to a meaning exposition.

The process of perception/understanding also has its own characteristics. A user finds and recognizes objects related to a MAc task, i.e. in a task context and available knowledge. Finding consists in identifying found objects with a certain sample – an existing knowledge “as a whole”, and recognition – in identifying individual “useful” properties identified in a text. At the search beginning, it is possible, albeit rather tentatively, to distinguish two “extreme” cognitive states:

1. The user knows a problem essence and has some idea for expected outcome, in which case the user examines documents for an expected image or its elements;
2. The user's knowledge is so incomplete or inconsistent that it does not allow forming any coherent solution image. In this case, the user needs to carry out analytic-synthetic actions over the text content in order to find “starting points”, to get a general idea about the nature and distribution of meaning elements (those and/or other characteristic concepts, chains of their use).

If the user has search objects idea (a separate concept or links between concepts), a “neighborhood analysis” and “path search” functions are used to localize a searched meaning.

Using the “neighborhood analysis” search mechanism, the user iteratively forms personal knowledge about elements that form the graph by sequentially traversing the vertices adjacent to an initial vertex. The initial vertex can be chosen by the user at random, or based on the potential importance of concept associated with the vertex.

The “path finding mechanism” use makes it possible to highlight individual facts related to given concepts. The application of different path selection algorithms (based on path length, on included in the path concepts weights sum, etc.) gives possible paths variety, which provides the basis for new knowledge synthesis. By analysing the different paths in the graph linking specified vertices, the user refines the knowledge about the connections between the concepts, or forms new ones, by comparing the selected paths with the available activity schemes.

Assuming that the concept importance is proportional to its use frequency (weight) in text, graph vertices geometric size can be set in proportion to the corresponding concept frequency. In particular, it would help to select as starting points those whose labels contain the most frequently used concepts that reflect the main text content.

In the prior knowledge absence, the user should be given the opportunity to represent the knowledge graph in the form most suitable for both the task to be solved and the perception. For this purpose, different algorithms are provided for stacking the graph vertices: based on path length, based on force method, based on functional model.

The use of search filters (vertex names and/or relationship classes, aspects) will ensure that the reduction space is manageable according to the cognitive situation.

Interactive use of graph representations makes it possible to highlight concepts relevant to the current problem and build a solution image (form its conceptual framework), and the transition from graph elements to corresponding text fragments makes it possible to form a “quoted” solution image from corresponding fragments of found texts.

That is, search operations on knowledge graphs combined with visualization tools become information search tools.

## Cognitive subject Tree as presentation tool OF Knowledge STRUCTURE AND CONTENT

For the complex search type [10], it is characteristic that in addition to retrieving the documents itself (that is the final aim in the case of factual and thematic search), the direction of research also has to be searched – paradigm (point of view, conditions, etc.), within which the information blocks of retrieved documents “will construct” a new knowledge.

But for all that, existing IRS operate at the syntax level, and in the best case, semantics, that can provide no more than a construction plausible issue (“possible world” from the point of view accepted in the system). But in science there are many “possible worlds”, as well as the system’s “point of view” may be different from the user's – a solver of a particular problem. Thus, it is necessary to “introduce” in the search process a pragmatic component, which will determine the cognitive position of the user – the conditions and circumstances of the possible solution. This involves using in the search process more complicated (than verbal expression) structures, that binds activity real objects (or more precisely, their descriptions), the system of concepts (operational objects of knowledge process) and sign systems (are used for describing objects) with taxonomy (tree of objectives) of the problem, defining the structure of the search process, and in essence – the direction and conditions of cognition process.

That is for personal information systems that are intended primarily for information-oriented user, an information search interactively-iterative process visualization and a consecutive formation of operational workspace becomes essentially significant, because in that case it is not limited to individual set of retrieved documents that are relevant to the current information need, but also as an interconnected (within the search direction) set of different nature objects and purposes – documents, terms, classification structures, search queries, etc.

To generate user-defined sight of the application domain in conjunction with the conventional system of concepts and classification, on the interface level it was proposed to use a specialized constructor [11], called Cognitive Subject Tree (CST). Its basic functional purpose – intensional (through the system of classifications) and extensional (in sets of documents, fragments of conceptual and terminological systems) represent individual knowledge correlated with the generally accepted.

CST is (generally) a network structure whose nodes contain references to documentary lexical filling of a separate rubric (theme node), and the arcs are oriented according to the main type of connection “part-whole”.

Separate rubric of the CST, considered as an independent object, includes two mutually complementary parts – to document references, that substantially represents thematic, and elements of linguistic support as a lexical-semantic basis.

Access to documents is provided either in a linear sequence (documents can thus be sorted according to the values of individual data elements), or clusters, that are usually thematically (or statistically) grouped.

Elements of the linguistic support are presented by the fragments of conceptual and terminological structures, which can be ordered semantically (classifications, thesauri, ontologies) or by frequency (dictionaries).

For the visualization of CST was designed a tree GUI-component, allowing the nodes state (content) visualization, and using of the standard tools for creating a new node-rubric and, particularly, structural modification based on technology “drag and drop”. It is possible to use structural operations – add/delete rubric (nodes) (with or without its descendants) at any level, “drag” rubric (with its descendants) to another branch of a tree, etc.

For analysis of the rubric current state a two-level scheme is proposed. On first level an integral indicator is used, which in accordance with the decomposition of the properties, graphically (using color coding) describe the facts of a separate rubric fullness with a specific content. The second level of visualization is designed for displaying the results of comparative analysis of the individual rubrics state to support a decision on further development or modification of the CST.

The possibility of quantitative analysis is provided by two factors.

1. The basis for the analysis is the combination of intensional and extensional representation of the investigated domain. Intensional component is represented by the structure of the CST (the system of division signs) essentially reflects the path of knowledge – the essential properties of the object of knowledge with point of view of the person (the researcher), and the priorities and order of extraction of the investigated components of the object. Thus extensional component represents the content of tree nodes, allows the use of quantitative methods for the comparative assessment of the separate directions of the cognitive process. Finally, the CST complexly represents the process, reflecting both the application domain (documentary representation) and the instrument of knowledge –conceptual system that are used.
2. Although the properties of the CST are different in nature (i.e., belongs to a different types of objects – documents, requests, word-lists, thesauri, subject indexes, ontologies) and are not directly comparable with each other, they all have the semiotic nature (the main significant element in information retrieval is the term – the word, classification code, phrase, etc.) that allows their comparison (of course, with the transformation, that, in some cases, can be ambiguous).

In the interface the second level is represented by a separate dialog windows containing diagrams and tables with (mostly) quantitative analysis of the content of the separate property of the rubrics, selected by user.

Technologically the most constructive way of such an analysis is to extract rubrics that have descendants, in order to determine the correctness of its division. In accordance with the accepted (theoretically) classification scheme of division (intersection of subclasses must be empty, the union of subclasses must give the divisible class, classification tree should be balanced) the requirements for verification the cardinality of intersection of rubrics was formulated.

If the verification of how complete the union of all created rubrics reflects the developing application domain is possible only through the involvement of the researcher, the verification of the subject index balance and the intersection of its rubrics are sufficiently formalized.

## Examples[[1]](#footnote-2)

### Aspect projection

Let's consider an example of a search on knowledge graphs on the topic “Proryv Project” using the aspect projection operation to build an aspect representation in the “efficiency” aspect.

As a result of an information retrieval on the topic “Proryv Project” using traditional search mechanisms, 7 relevant documents were found, which were combined into a single text. According to this text, using xIRBIS [9], an ontology graph was formed[[2]](#footnote-3).

The “aspect projection” function was applied to the ontology graph in the “efficiency” aspect, as a result an aspect representation was formed – the subgraph shown in Fig. 1, which includes graph elements that satisfy the conditions for setting the aspect, including the structural-linguistic relationships “context”, “context-thesaurus”, “equals”, “contains an entity”, “circumstance of use”.

 FIG. 1. Aspect representation in the aspect of “efficiency” on the topic “Proryv Project”

In the subgraph, the search for vertices was performed using the expression “Proryv Project”. Further, when visually examining the neighborhood of the found vertex (Fig. 1, top side), one can detect a “Russian export development” vertex (Fig. 1, lower side). Let's build a path between vertices using the pathfinding function, specifying “Proryv Project” vertex as a source vertex, and “Russian export development” as a target vertex. The presence of such a path allows to assume[[3]](#footnote-4) that the possible effect of the implementation of the Proryv Project is the development of Russian exports.

The transition from the vertices of the considered path to the corresponding fragments of the source text allows to confirm this statement, since the source text contains the following fragments[[4]](#footnote-5): *“The* ***Proryv Project*** *implemented by the State Atomic Energy Corporation ROSATOM is aimed at achieving a new quality of nuclear energy… The purpose of activities under the* ***Proryv Project*** *is the creation of* ***nuclear-energy complexes*** *that include NPP, spent nuclear fuel reprocessing plant and fuel re-fabrication plant… The* ***Pilot-demonstration energy complex*** *(PDEC) is being built at the Siberian Chemical Combine site as part of a* ***power unit*** *with a* ***lead-cooled BREST-OD-300 reactor*** *and an on-site closed nuclear fuel cycle facility, which includes a* ***plant for reprocessing*** *(PR) irradiated mixed uranium-plutonium (nitride) fuel and a fabrication/refabrication plant (FRP) for the manufactures initial startup fuel assemblies from offsite materials and, eventually, nuclear fuel from reprocessed SNF products. The construction of the* ***BREST-OD-300 fast neutron reactor*** *should lead to the* ***Russian exports development*** *in the nuclear industry...”*

### Semantic scaling

In general, as knowledge semantic scaling (granulation) we will mean an information amount change about objects, integrally provided for perception. Reducing information within the graph-theoretic model framework is possible due to (a) decreasing / highlighting / selecting in a cognitive activity field a reality fragment described by a graph and, accordingly, displayed vertices and relationships number (that is, a subgraph highlighting / formation), and / or (b) decreasing in cognitive activity field available objects analyzed properties, that is, by transition to higher level concepts and relationships.

Scaling types are determined by features nature (essence): will be used in operations objects-instances or object properties. In the first case – it`s enlargement as objects aggregation on an inclusion basis (part-whole relationship). In the second case – it`s generalization by bringing objects-concepts/relationships (constituting a generalized image) to a generality higher level (genus-species relationship).

In a knowledge graph context, enlargement is an operation in which in routes where all relationships belong to the “part-whole” class, original graph intermediate vertices (i.e., non-initial and non-final) are “hidden”. Generalization is an operation in which more specific concepts are replaced by generic ones (i.e., objects common properties are used) while preserving original relationships.

Semantic enlargement of knowledge graph constructed from text is performed by identifying arcs on the graph that correspond to “part-whole” type relationships, and replacing arc with incident vertices with a vertex “whole” while preserving all other arcs of both vertices. Thus, the subgraphs, which presents dividing the whole into parts result, are replaced by one vertex.

Revealing relations “part-whole” is ensured by using relationships classification [12] in knowledge graph constructing. Top level of this classification is represented by combination of three facets, reflecting the separate and aggregate ratio (individual – individual, individual – whole, all parts – whole), the reality / model ratio, and also reflecting a relationship manifestation form. To encode classes, a positional coding system is used, the first code position is reserved for the ratio between separate and aggregate facet. Thus, the relationships class code included in a hierarchy, growing from the top-level class with facet focus, representing the relationship “individual – whole” (the code of this class code begins with numeral “2”).

Fig. 2 shows the graph obtained as the enlargement operation application result.



FIG. 2. The graph obtained as result of the enlargement

The fig. 2 shows that “pump” is the most significant concept, which represents the entity of which the “removable part” is a part. This is obtained as an absorption result of vertices connected with the “pump” vertex by the “part–whole” relationships. In this case, the number of vertices is reduced.

However, it should be noted that generalization (like other intellective operations) cannot be performed by a separate action on a separate concept. A concept (more precisely, its specific content - meaning) depends on surrounding concepts (as context) and, in turn, determines the meaning of others, thereby creating a semantic integrity. Similarly, relationships identification, that reflects reality interrelations and properties change, depends on context. That is, first it is necessary to reveal and identify a base essence of text content and then – to determine each entity significance and relationships class (from role point of view in scaled image being formed). Based on functional-cybernetic model, the essence may be represented: (a) by a predominant properties class (area) presented in text; and (b) by a predominant text aspect orientation and / or by a nature of goal effect.

Accordingly, in knowledge graphs scaling operations all concepts and relationships should be analyzed and evaluated, and not in a separate aspect. And the scaling procedure should be based on both of the above approaches, and with the declarative means use – concepts and relationships thesauri, and also use a top-level ontology.

## CONCLUSION

Semantic documentary information search focused on cognitive processes can be considered as a supporting activity, the result of which replaces a main creative activity of a person. This determines that such a process includes the following dialectically complementary actions:

1. selection of formally relevant documents based on queries from an information array and deep semantic search for relevant semantic fragments within them on the basis of constructiveness in relation to a problem being solved;
2. formation of a plausible conceptual image of a hypothetical solution by means of operations on graphs constructed on the basis of facts extracted from found texts;
3. synthesis of a new quasi-text – the linking of relevant fragments-“quotations” ordered in accordance with a conceptual image “in the sense”;
4. systematization and evaluation of a received content – “fitting and embedding” a new meaning into an accepted system of knowledge.

In such search process, not only a conceptual (ontological) image of solving user's pragmatic task is formed, but also a development of a language – a research tool and a means of communication – takes place (and is fixed). A knowledge graph is essentially a reflexive image of an existing solution/state of a problem situation, but since reflection is a mapping of a text content presented by the conceptual graph to the subject's problem situation, such a graph also sets some vision of the future solution. A knowledge graph can be a “polygon with contour maps” where user implements a trajectory of both informational and subject search: understanding of using key concepts expediency is provided through these concepts visualized contexts and paths construction between them.

The given examples of semantic scaling operations (similar to abstraction operation in cognition) show that their use reduces the number of displayed vertices, facilitating the perception of the graph without significantly losing the meaning of describing the pump as a set of elements that make it up.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (state assignment project No. 0723-2020-0036).

References

1. Dahlgren, K., “Demonstration of Cognition Search”, Semantic Computing (Proc. Int. Conf. Pittsburgh, USA, 2010) IEEE Computer Society, Washington, DC, US (2010) 454–455.
2. What is “cognitive search” in Azure Search?, https://docs.microsoft.com/en-us/azure/search/cognitive-search-concept-intro, last accessed 2022/01/12.
3. Microsoft Azure official YouTube channel, How to build AI applications with Cognitive Search, <https://www.youtube.com/watch?v=k5xScEyyI4M>, last accessed 2022/01/12.
4. Gualtieri, M., The Forrester Wave: Cognitive Search and Knowledge Discovery Solution, report, Forrester research, 2017, https://go.forrester.com/blogs/17-06-12-cognitive\_search\_is\_the\_ai\_version\_of\_enterprise\_search/, last accessed 2022/01/12.
5. Bast, H., Buchhold, B., Haussmann, E., Semantic search on text and knowledge bases, Foundations and Trends in Information Retrieval 10 (2–3) (2016) 119–271.
6. Hogan, A. et al., “(Re)Defining Knowledge Graphs”, Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web: Report from Dagstuhl Seminar September 9–14, 2018.
7. Maksimov, N. V. et al., “Graph-Ontology Model of Cognitive-Similar Information Retrieval (on the Requirements Tracing Task Example)” (Biologically Inspired Cognitive Architectures Meeting, Natal, Brazil, 2020), Springer, Cham (2021).
8. Maksimov, N. et al., Knowledge representation models and cognitive search support tools, Procedia Computer Science 169 (2020) 81-89.
9. Maksimov, N. V. et al., Dokumental'naya informacionno-analiticheskaya sistema xIRBIS (redakciya 6.0), Svidetel'stvo ob ofitsial'noi registratsii programm dlya EVM (Document Information-Analytical System xIRBIS (edition). The Certificate on Official Registration of the Computer Program), No. 2020661683 (2020). (in Russian)
10. Golitsyna, O.L., Maksimov, N.V., Popov, I.I., Informatsionnie Sistemy: Ucheb. Posobie (Information Systems: Training Manual), Forum, Moscow (2007). (in Russian)
11. Maksimov, N.V., Golitsyna, O.L., The architecture and softwar-informational means for cognitive processes support. Rep. at the conf. Information Technology in Education – 2009. (in Russian)
12. Maksimov, N. V. et al., Systematization and identification of semantic relations in ontologies for scientific and technical subject areas, Automatic Documentation and Mathematical Linguistics 52(6) (2018) 306–317.
1. In all examples, graphs are built according to texts in Russian, then labels of vertices and arcs are translated into English using the INIS Multilingual Thesaurus. The graphs in vertices contains named entities (concepts, names, values, etc.), and in arcs – typed relationships, extracted from the text also taking into account their location. Numeric values in arc labels contain the relationship class code from the taxonomy of functional relationship classes [12]. [↑](#footnote-ref-2)
2. Due to a large volume of graphs (thousands vertices and arcs), only fragments of graphs are given. [↑](#footnote-ref-3)
3. It should be noted that the developed technologies are aimed at solving information retrieval problems and do not belong to the class of expert systems. The system provides the user with materials for selection and analysis, in particular, information about the possible existence of a relationship between objects, and does not provide a solution - reliable cause-and-effect relationships, and does not perform logical conclusions. [↑](#footnote-ref-4)
4. The words and phrases corresponding to the graph elements are highlighted in bold. [↑](#footnote-ref-5)